Jiayin Gao* and Martine Mazaudon

Flexibility and evolution of cue weighting after a tonal split: an experimental field study on Tamang

https://doi.org/10.1515/lingvan-2021-0085
Received May 31, 2021; accepted June 15, 2021; published online June 17, 2022

Abstract: We conducted a perception experiment in the field to examine the synchronic consequences of a tonal split in Risiangku Tamang (Tibeto-Burman). Proto-Tamang was a two-tone language with three series of plosives and two series of continuants. The merger of its continuants provoked a split of the original two tones into four, two high and two low, which combine pitch and phonation features. The quasi-merger of the voiced and voiceless plosives left sporadic remnants of initial plosive voicing in low tone syllables. A previous production study has shown that speakers use pitch and phonation features concomitantly to distinguish high from low tones, while producing initial plosive voicing only marginally with low tones. The present perception study establishes the preeminence of the pitch cue, but also confirms the effective use of the two older cues in tone identification. An apparent-time analysis shows the phonation cue to be less used by younger speakers, in keeping with the historical evolution. The use of the residual voicing of plosives, instead of decreasing with younger speakers, is shown to increase. This result could be explained by an increased contact of the young generation with Nepali, a toneless Indo-Aryan language with a four-way initial plosive contrast.

Keywords: breathy voice; cue weighting; phonation; Tibeto-Burman; tonogenesis

1 Introduction

1.1 Tamang and the linguistic situation in Nepal

Nepal is home to around 130 languages of Indo-Aryan, Tibeto-Burman, and other families, some tonal, most toneless. With over 1,300,000 mother-tongue speakers (Nepal census 2011), the Tamang community constitutes the largest non-Indo-Aryan group in Nepal. Until the mid-twentieth century, the Tamang, mostly endogamous and village-dwellers, were largely monolingual. A few Tamang were literate in Nepali, the Indo-Aryan national language, or in Tibetan. Very occasionally, Tamang was jotted down for private use in one of the corresponding alphabets, Devanagari or Tibetan. However, in the last 40 years, bilingualism and schooling in Nepali have become widespread.

Tamang belongs to the Tamangish group (also known as TGTM or Tamang-Gurung-Thakali-Manangke), a branch of the Bodish section of Tibeto-Burman which also contains Tibetan, as well as Tshangla (ISO [tsj]) and Kurtöp (ISO [xkz]) in Bhutan. There are several dialects of Tamang. The current study bears on Risiangku Tamang, an Eastern Tamang variety (ISO [taj]). The administrative unit of Risiangku (alternative names: Lisankhu, Risingo; 27°6′ N, 85°8′ E) has 3,700 residents, 2,300 of whom are Tamang (Nepal census 2011). Risiangku is recognized by the Tamang as a cultural center, boasting a Buddhist temple since AD 1093 (Lama 1959).

***Corresponding author: Jiayin Gao**, Laboratoire de Phonétiquet et Phonologie, Langues et Civilisations à Tradition Orale, LabEx – Empirical Foundations of Linguistics, Paris, France; and University of Edinburgh, Edinburgh, UK, E-mail: jiayin.gao@ed.ac.uk. https://orcid.org/0000-0001-7572-4482
Martine Mazaudon, LabEx – Empirical Foundations of Linguistics, Paris, France; and Centre National de la Recherche Scientifique, Paris, France, E-mail: martine.mazaudon@cnrs.fr

1.2 The Tamang tonal split in an areal context

All modern TGTM languages are tonal, with systems of four tones defined on lexemes, which are largely monosyllabic (almost all verb roots and half the nouns). Suffixes have no independent tones and the tonal characteristics of the initial morpheme extend over the whole phonological word, constituting a word-tone system, first described in Pike (1970) and Mazaudon (1973). Proto-TGTM, the ancestor of all TGTM languages, has been reconstructed with three plosive series (voiceless aspirated, voiceless unaspirated, and prevoiced), two series of continuants (voiceless and voiced fricatives and sonorants), and only two tones (Mazaudon 1978, 2012). The gradual loss of the voicing distinction on initial consonants was the source of the split of the old two-tone system into the modern four-tone systems. While some residual voicing of initial plosives survives in some of the languages, the two series of continuants have completely merged in all TGTM languages, thus finalizing the phonemicization of the new tonal contrast.[1]

The consonantal mutation and tonal split which we presently observe in Tamang is an instantiation of the vast Asian mutation which started more than a millennium ago in Middle Chinese and swept through most parts of Asia in various forms (Maspero 1912; Haudricourt 1961, 1965). Schematically, the loss of a voicing contrast on initial consonants was replaced by two orders of contrasts: (1) in previously tonal languages, like Middle Chinese, Vietnamese, or Thai, the number of tones was doubled or sometimes tripled; (2) in previously toneless languages, like Khmer, a register contrast was created, usually manifested by a doubling of vowel timbres, often associated with secondary features of breathiness, pitch, or vowel length (Ferlus 1979; Huffman 1976).

Multiple cues are known to be involved in register languages, and they have been studied experimentally (for a review see Brunelle and Kirby 2016). But tonal languages are not pure pitch either. Haudricourt (1965) proposed that at the beginning of a tonal split, breathy voice often accompanied the devoicing of the old voiced plosives. Breathiness was retained alongside some voicing of initial consonants in the Wu dialects of Chinese (for recent experimental studies, see, e.g., Gao 2015; Tian and Kuang 2021; Jiang et al. 2020; Zhang and Yan 2018).

This stage can be observed in the Tamangish languages, where multiple cues contribute to the tonal contrasts but their realizations and relative weights differ between varieties. In conservative varieties, there is a clear high versus low tonal contrast: the two high tones, produced with high pitch onset and modal voice, co-occur with aspirated and unaspirated voiceless plosives, while the two low tones, produced with low F0 and breathy voice, co-occur with a single series of plosives, unaspirated, occasionally realized with partial or full voicing. In evolved varieties, such as Manangke, Marphali, or Taglung Tamang, further changes have obscured the etymological picture (Hildebrandt 2005; Gao and Mazaudon 2017; Mazaudon 1978).

1.3 Summary of production data in Risiangku Tamang

To better understand the synchronic relationships between features that arose from the tonal split, an earlier experimental investigation was conducted on one of the conservative varieties, Risiangku Tamang, using electroglottographic data from five male speakers in their 30s–40s (Mazaudon and Michaud 2008; Michaud and Mazaudon 2006). The phonetic description of the four tones is summarized as follows:

(1) F0 height and contour: T1 is the highest and has a falling contour; T2 is the second highest; T3 is low and rising; T4 is the lowest and falling.
(2) Phonation: T1 and T2 are modal; T3 and T4 are breathy (or "whispery" as used in that study) but T4 to a lesser degree. On average, glottal open quotient is above 50% for T1 and T2 for all speakers, and higher

1 Although the link between voiced plosive and low tone is and has been the most salient connection in the tonogenesis literature to this day, as early as 1949, and more clearly in 1961, Haudricourt remarked, on the evidence of the Tho (Tai) dialect of Cao Bằng, that tonogenesis/tonal split could rest on the merger of the continuant series, in the absence of the devoicing of the plosives (Haudricourt 1949, 1961; Pittayaporn and Kirby 2016).

(from 55 to 71%) for T3 and T4, which suggests that the phonation difference is between modal and breathy, rather than, say, laryngealized/tense and lax (Maddieson and Ladefoged 1985).

(3) Voicing: 20–30% of the initial plosives of low tone words are fully or partially voiced, as opposed to almost 0% in the high tone words.[2]

In the Tamang word-tone system, the F0 contour extends over the entire word but noninitial syllables are subject to intonational variability, as shown in Figure 1, which plots the mean F0 curves in semitones of a limited number of words (n = 43) produced by the male participants of the current study. Figure 2 shows the

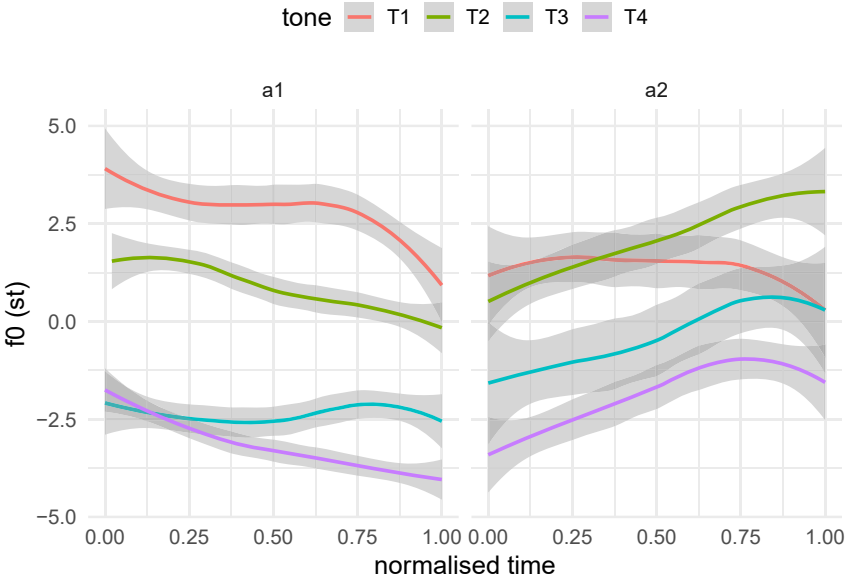

Figure 1: Mean smoothed F0 curves in semitones of the four tones produced by the male participants of the current study: disyllabic /**pa(:)-pa**/. Shading indicates 95% confidence interval.

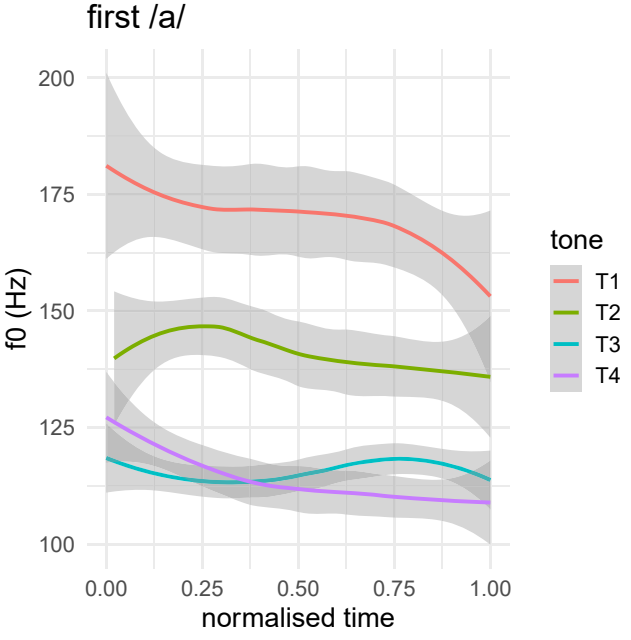

Figure 2: Mean smoothed curves in hertz of the four tones produced by the male participants of the current study: first /a/ of /**pa(:)-pa**/. Shading indicates 95% confidence interval.

2 As in other Sino-Tibetan languages, syllable-final plosives are voiceless in Tamang. In word-internal position, unaspirated plosives are fully voiced, except when following a preceding syllable-final plosive.

mean F0 curves in hertz of the first syllable alone. The F0 curves are similar to the plots in Michaud and Mazaudon (2006) and Mazaudon and Michaud (2008), which can be consulted for more details, as well as Mazaudon (1973) and Mazaudon (2005: 84, Figure 4).

1.4 The current study: perception of newly emerged and residual cues

Based on previous production data, the current study investigates the equilibrium of plosive voicing, phonation, and pitch height in the *perception* of high versus low tones in Risiangku Tamang. In early 2017, we conducted a tone identification experiment in Nepal with native speakers of Risiangku Tamang, adapting as much as possible a laboratory setting to the field, with the following questions in mind:

- Pitch and phonation are used in signaling the tonal contrast in production. Are they equally important in perception, and how do they interact?
- Plosive voicing is marginally produced and is highly redundant. What, then, is its role in perception, and how does it interact with pitch and phonation?
- How do these features evolve after the phonemicization of the new tonal contrast? Will Risiangku Tamang necessarily follow the path to a more evolved variety, in which pitch becomes the primary cue? Can we observe evidence for such an evolution in apparent-time data?

2 Method

2.1 Base words

There are almost no quadruplets with solely a tone difference. We selected a quasi-quadruplet relatively easy to synthesize: /[1]**pa-pa**/ 'to be too liquid', /[2]**pa**ː**-pa**/ 'to be rough (taste)', /[3]**pa-pa**/ 'to bring', /[4]**pa**ː**-pa**/ 'to pile up' (with numbers representing tones from highest to lowest). (The phonetic realization of the second /p/ in each word is voiced, as explained in note 2, Section 1.3.) Each word is a monomorphemic stem followed by an affix. They are common words, but /[4]**pa**ː**-pa**/ 'to pile up' is less familiar to young speakers. All stimuli were synthesized with equalized segmental duration so as to minimize the effect of vowel length.

2.2 Stimuli

Stimuli were synthesized with target words preceded by a deictic /[2]**tsu**/ 'this', using an articulatory synthesizer, VocalTractLab 2.1 (Birkholz 2013), followed by other modifications detailed below. The target words were $C_1\,V_1\,C_2\,V_2$ sequences, where C_1 and C_2 were both labial plosives, and V_1 and V_2 were /a/. Five parameters were manipulated on the target word. Examples of spectrograms are given in Appendix A.2 of Supplementary material. All stimuli can be found in the Supplementary material.[3]

(1) Degrees of breathiness of $C_1\,V_1$: labeled here as "modal", "breathy", and "super breathy (sup_br)". Three degrees of glottal opening were synthesized in VocalTractLab 2.1 to simulate three degrees of breathiness, by modifying the upper and lower rest displacement and the arytenoid area (see gestural scores in the Supplementary material). Their difference in breathiness was confirmed by H1–H2 measure over the target vowel: 2.5, 7, and 12 dB, for modal, breathy, and super breathy, respectively. Segmental durations and intensity contours were then equalized across items.

3 The Appendices and the Supplementary material are offered in the online version of the article, and also in our Open Science Framework page: https://doi.org/10.17605/OSF.IO/9RUD6.

(2) Presence or absence of the prevoicing of C_1: VOT at –70 or 12 ms. For each unvoiced stimulus, we created a version with prevoiced C_1 with a voice lead of 70 ms computed as a low-pass filtered and attenuated extract of the vowel following a voiceless plosive.

The resulting sequences were then manipulated for F0 with TD-PSOLA (Valbret et al. 1992) implemented in Praat (Boersma and Weenink 2016) (F0 was at 115 Hz steadily on the vowel in /^{2}**tsu**/):

(3) F0 onset of V_1: at 115, 130, 145, or 160 Hz.
(4) F0 slope of V_1: linear rise of 10 Hz, versus fall of 20 Hz.
(5) F0 slope of V_2: linear rise of 5 Hz, versus fall of 15 Hz. The F0 onset of V_2 is always 5 Hz lower than the F0 offset of V_1.

The endpoints of F0 onset were based on the production of speaker M3 in Mazaudon and Michaud (2008) (see Michaud and Mazaudon 2006: Figure 1), which are comparable to the F0 production in our data in Figure 2. The breathy and super breathy versions were created after trials of different parameters to meet the auditory judgment of breathiness of the first author, a trained phonetician. This made a total of 96 stimuli (3 degrees of breathiness × 2 voicing parameters × 4 F0 onsets × 2 F0 slopes of V1 × 2 F0 slopes of V2).

2.3 Participants and data collection

The results of 28 participants (14 female, 14 male) of the experiment, aged 33–79 (mean = 49, SD = 12.3) at the time of the experiment, will be reported.[4] All are bilingual speakers of Risiangku Tamang and Nepali. We discarded the data of five additional participants who were unable to complete the task.

The experiment was conducted using PsychoPy 2 (Peirce and MacAskill 2018). Participants were tested individually in a recording studio in Kathmandu, or a quiet room in the village of Risiangku. They wore a Sennheiser HD518 headphone and responded on a laptop computer. In each trial, an auditory stimulus was presented in two repetitions, and the participant was invited to select the most appropriate among four pictures displayed in the four corners of the screen (see Appendix A.1 in Supplementary material). The response time-out was set to 10 s, in consideration of the unfamiliarity of the participants with this type of task. Participants took two short breaks during the experiment session. Stimuli were presented in a different randomized order for each participant.

The experiment session was preceded by two training sessions during which participants were presented with natural stimuli produced by two male speakers. The first training session of 11 trials aimed at familiarizing all the participants with the task, using different items from the target words. The second training session of 10 trials was designed to make sure that participants could identify each naturally produced target word from the quasi-quadruplet and associate it with the intended meaning represented by the picture. Feedback was given for correctness with a happy or a sad emoticon. Participants who had difficulties in understanding the instructions repeated the failed session once or twice. For most participants, we recorded their production of the four test words for qualitative observations.[5]

2.4 Limitations

Since only one quadruplet was used, our results are limited to one place of articulation (bilabial) followed by the vowel /a/. It is possible that the perception of VOT and/or F0 is affected by place of articulation (Peralta 2018) and/or the following vowel. The reason why we refrained from including more words or repetitions was

4 The study was approved by Tamang Nhankhor (the Tamang association for the preservation and promotion of the Tamang language and culture), and included a waiver of informed consent as it was determined that the research presents no more than minimal risk to subjects and that a waiver of informed consent would not adversely affect the rights and welfare of subjects.
5 See the Open Science Framework page: https://doi.org/10.17605/OSF.IO/9RUD6.

to avoid imposing a long attention span on our participants, who are mostly unaccustomed to computers and sometimes have difficulties understanding the instructions. As it was, most villager participants spent more than 30 min in completing all the sessions. Consequently, the limited number of trials may result in failure to detect an existing statistical significance due to a low statistical power (e.g., Kirby and Sonderegger 2018). However, we think that an increase in the number of trials could increase error rates, which would be undesirable. Another drawback of our study is that we were unable to establish a sociological profile for each participant. We will examine the age factor in the following, but other sociological factors were not controlled for.

3 Results: synthesized stimuli

The data set and the code to reproduce the plots and analysis are given in the Supplementary material. Results of the identification of natural stimuli in the second training session are reported in Appendix B of Supplementary material. Here, we report the results of synthesized stimuli in the experiment session.

3.1 Identification rates and response times

We first tried exploring how listeners identify the four tones. Possibly due to the equalized vowel length and the oversimplified F0 slope manipulation in our stimuli, the results were inconclusive, except that the ratio between tone 3 and tone 4 responses increased as the degree of breathiness increased (modal: 0.87; breathy: 1.11; super breathy: 1.46), in line with the production results in Mazaudon and Michaud (2008). Since we were mostly interested in the perceptual difference between high and low tones, in the following analyses, we group tones 1 and 2 into high tone responses, and tones 3 and 4 into low tone responses. Also, we do not expect F0 slopes of V1 or V2 to interfere in the identification of high versus low tones. Note that although each stimulus was presented only once to the participants, the averaging of responses over F0 slopes means that there were four tokens for each F0 + PHONATION + PREVOICING condition.

Figure 3 shows the averaged identification rate of high tones. In the following, we provide descriptive statistics as well as the results from a generalized linear mixed model fitted to the response data (using the `lmertest` package [Kuznetsova et al. 2017] in R [R Core Team 2019]; formula and outputs in Appendix C.1 of Supplementary material). As expected, high tone identification rate increases as F0 onset increases ($14\% < 33\% < 58\% < 71\%$, $\beta = 3.85$, $\mathrm{SE} = 0.18, z = 21.28, p < 0.0001$), and as breathiness decreases

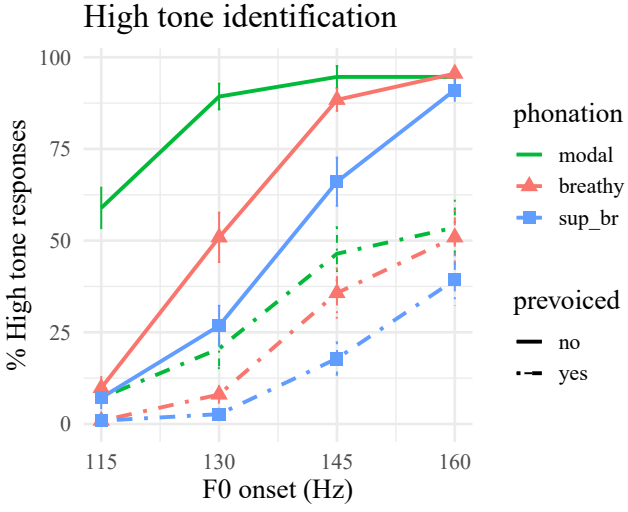

Figure 3: High tone identification curves (T1/T2) by F0 onset, phonation, and prevoicing. Error bars represent one standard error.

(31% < 43% < 58%, $\beta = -1.68$, SE = 0.12, $z = -13.73$, $p < 0.0001$). Unlike the marginality of prevoicing in the production of low tones, prevoicing leads to a strong bias towards low tone perception (24% for unvoiced vs. 64% for prevoiced stimuli, $\beta = -3.67$, SE = 0.34, $z = -10.73$, $p < 0.0001$). The model also shows that AGE interacts with two main factors: the effect of PREVOICING decreases with age ($\beta = 0.05$, SE = 0.03, $z = 2.05$, $p < 0.05$), while the effect of PHONATION increases with age ($\beta = -0.03$, SE = 0.01, $z = -3.68$, $p < 0.0005$).

Two asymmetries can be observed. First, at the highest F0 onset, prevoiced stimuli reduce high tone responses almost by half, regardless of phonation, while unvoiced stimuli do not have the opposite effect at the lowest F0 onset. Second, for unvoiced stimuli, modal voice leads to a strong bias towards high tone identification, including at the lowest F0 onset, while breathy and super breathy stimuli do not bias toward low tone identification at the highest F0 onset. (See Appendix C.1 in Supplementary material for results of pairwise comparisons, using the `emmeans` package [Lenth 2019].) Thus, the prevoiced property conflicts with high tone identification, and modal voice with unvoiced plosive conflicts with low tone identification.

Response times are also affected by the (in)congruence of cues. Figure 4 shows that response time is the lowest when all three cues are congruent: (1) breathy voice + prevoiced + lowest F0, or (2) modal voice + unvoiced + highest F0. Response times increase when two cues are in conflict, where the identification rate is around 50%: (1) prevoiced + highest F0, or (2) modal voice + lowest F0. (See Appendix C.2 in Supplementary material for statistical models.)

3.2 Classification tree analysis

A classification tree analysis (CART) was conducted to further assess the relative importance of each cue in classifying high versus low tone categories, using the `rpart` package (Therneau et al. 2019) in R. Note that this data classification method divides the data into two maximally homogeneous groups at each split, but does not say anything about how a listener makes decisions about a tone category.

Figure 5 shows the classification tree, plotted by the `rpart.plot` package (Milborrow 2019) in R. F0 is used in the first split, with 138 Hz as cut-off. When F0 is above 138 Hz, PREVOICING is the best separator between the two response categories. When F0 is below 138 Hz, the prevoiced stimuli predict a low tone response, and unvoiced stimuli with a breathy voice also predict a low tone response. The asymmetries observed in Figure 3 are reflected here again to some degree. First, across the board, prevoiced stimuli predict a low tone response, while unvoiced stimuli have less powerful predictability. Second, for unvoiced stimuli with lower F0, modal

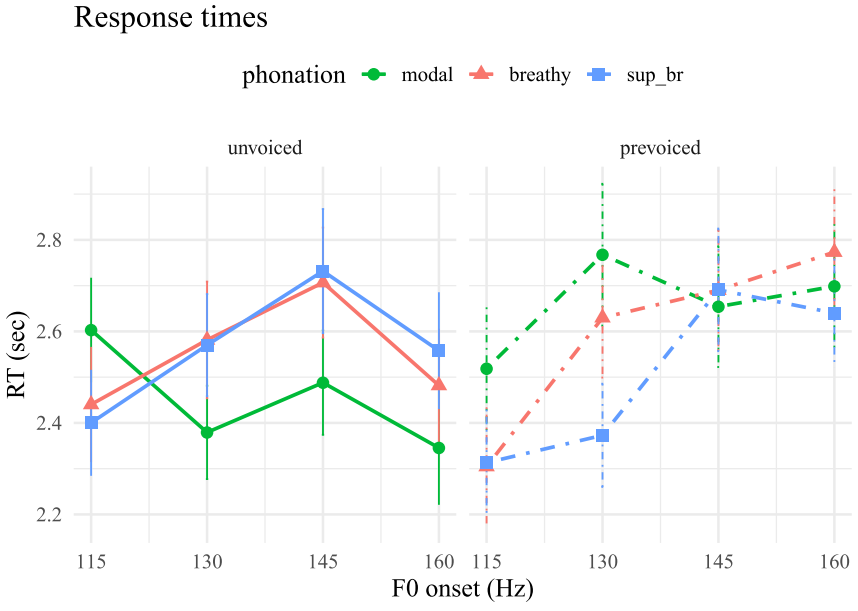

Figure 4: Response time curves by F0 onset, phonation, and prevoicing. Error bars represent one standard error.

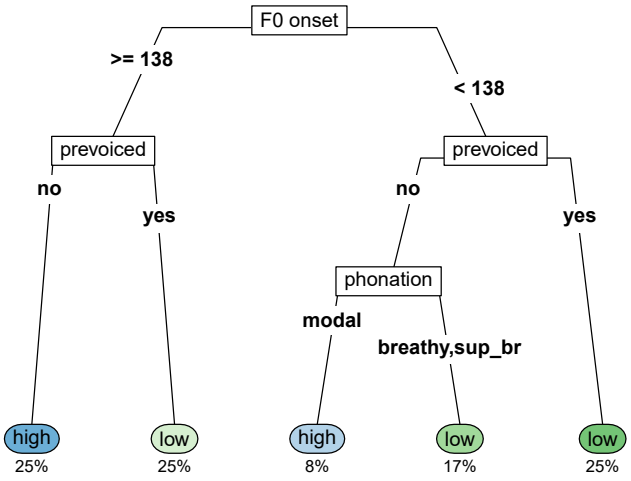

Figure 5: Classification tree of high versus low tone response categories. Below each node, the percentage of observations of each response category in the node is given.

voice predicts a high tone response, whereas for higher F0, phonation plays a smaller role in the prediction of the tone response.

3.3 Apparent-time variation

For each factor, high tone response differential was calculated for each participant by subtracting the minimum value (i.e., responses to stimuli with lowest F0 onset, or prevoicing, or super breathy phonation) from the maximum value (i.e., responses to stimuli with highest F0 onset, or without prevoicing, or with modal phonation). Higher response differential is taken as an indication of a greater use of a given cue by a given participant (see, e.g., difference score used in Idemaru et al. 2012). As shown in Figure 6, age correlates negatively with the response differential based on prevoicing, and positively with the response differential based on phonation, in line with the results of the generalized linear mixed model reported in Section 3.1. This suggests an increase in the use of prevoicing and a decrease in the use of phonation with decreasing age (for further analysis, see Appendix C.3 in Supplementary material).

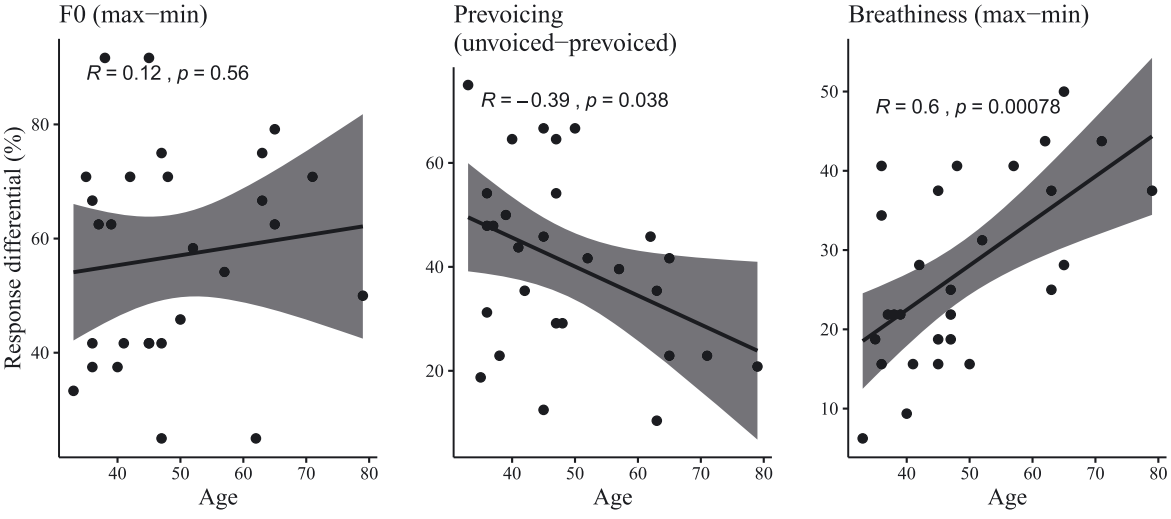

Figure 6: Scatterplot of high tone response differentials against age with fitted regression lines and coefficients: *left*, for F0; *center*, for prevoicing; *right*, for phonation.

4 Summary of results and discussion

Our perception study shows that speakers of Risiangku Tamang use pitch height, phonation, and plosive voicing in the identification of high versus low tones. Along with previous production studies, our empirical data lends support to the phonological characterization of tones in TGTM by a bundle of features. Some new patterns emerge from our perception data.

(a) At the group level, *pitch height* appears as the dominant cue, in keeping with its status as the new phonological feature.
(b) The *importance of the two secondary cues* of phonation and plosive voicing is demonstrated by the drastic fall in identification rate of high-pitched stimuli as high tone, and of low pitched stimuli as low tone in the presence of a *conflicting secondary cue*, as well as by the increased response time, which manifests the unease of the listener in such situations.
(c) A heavily breathy voice quality reinforces low tone perception but does not preclude high tone perception, while modal voice prevents low tone perception in more than 50% of cases. This suggests *breathy phonation as the neutral, or unmarked phonation* in Risiangku Tamang. Modal voice on the contrary seems strongly associated with high tones.
(d) Plosive voicing, which is only occasionally realized in production, takes a prominent place in low tone identification. *Its weight in perception is thus greater than in production.*
(e) The apparent-time analysis tends to show the expected decrease of the significance of the older feature of breathiness with younger speakers, but surprisingly shows an increase of the significance of the oldest feature, the de-phonologized plosive voicing.
(f) At the individual level, for most participants, pitch height has the highest perceptual weight, while for many others, prevoicing has the highest perceptual weight. This again suggests the variability and flexibility in the usage of multiple features (see analysis in Appendix C.3 of Supplementary material).

4.1 Breathy voice as unmarked

The strong conflict between modal voice and low tone identification raises a question: why should phonation matter more when associated with low F0 than with high F0? This has no obvious psychoacoustic motivation. No such asymmetry has been found in the perception of sine-wave overtones (Kuang and Liberman 2018: Figure 4). As a comparison, in a study on Shanghai Wu with similar stimuli settings, it has been found that the effect of phonation is the smallest at the lowest F0 onset (Gao et al. 2020: step 7 in Figure 5), but increases as F0 onset increases.

Breathy phonation is usually interpreted historically as a step on the way to the devoicing of initial voiced plosives. But in modern TGTM languages, breathy voice is present with low tones across the board, including on words with sonorant initials which were and remained voiced. In the case of sonorants – as opposed to plosives – the modified series was not the voiced series, but the voiceless series, which developed high pitch accompanied with modal phonation as they became voiced. This, coupled with the strong conflict between modal voice and low F0, suggests that a slightly breathy voice might be the neutral setting for Risiangku Tamang, while modal voice is "marked". A detailed study of phonation after continuant initials active in a tonal split would be instructive.

4.2 Perception-production delinking: the fate of plosive voicing

We found that the weight of plosive voicing is greater in perception than in production in Risiangku Tamang. Does this constitute an argument in the debate on production-led versus perception-led sound change (for a review, see Schertz and Clare 2020)? We think not.

Previous studies on incipient tonogenesis (e.g., Coetzee et al. 2018 on Afrikaans) or obstruent devoicing (e.g., Pinget et al. 2020 on Dutch; Gao and Arai 2019 and Gao et al. 2019 on Japanese; and Brunelle et al. 2020 on Chru) have shown that plosive voicing remains an important cue in perception while being variably used in production. These studies may be taken as arguments against the strong version of Ohala's (1981) hypothesis that misperception, taken in a broad sense, is the source of sound change.

However, the Tamang situation differs in several respects from languages undergoing incipient tonogenesis like Afrikaans (Coetzee et al. 2018) and Central Malagasy (Howe 2017). First, we are dealing with a tonal split, not the emergence of tone in a previously toneless language. The presence of tones in the protolanguage was plausibly a favoring context for the reinterpretation of coarticulation features issued from laryngeal characteristics of the initial consonants as tones. Second, the protolanguage had two series of continuants, unlike, for example, Afrikaans (Coetzee 2017). Other Tibeto-Burman languages that are undergoing tonogenesis – the creation of primary tones (e.g., Kurtöp: Michailovsky and Mazaudon 1994; Hyslop 2009; Peralta 2018) – or a tonal split (e.g., Lalo: Yang et al. 2015), although more similar to the Tamang situation in some respects, evidence a much lower degree of devoicing of their initial plosives.

Thus, Risiangku Tamang, with its much higher rate of devoicing, and situated after the phonemicization of the new tonal system due to the merger of two series of proto-continuants, is further advanced on the time scale of a change. Evidence from Risiangku Tamang does not bear on an incipient change but may be considered consistent with the reframing of Ohala's hypothesis by Kuang and Liberman (2018) and Pinget et al. (2020) along a time dimension in which production leads perception in the later stages of a change.

In fact, the situation may no longer reflect an ongoing change: after the completion of the change, listeners still attend to residual features as long as those features participate in the phonological categorization. Although plosive voicing is only marginally produced in Risiangku Tamang, when it is present, it cues low tones exclusively (since it never co-occurs with high tones). In this sense, the perceptual reliability of plosive voicing is greater than that of the other two features of pitch and phonation, which vary on a continuous scale.

4.3 Apparent-time variation, evolution, and language contact

Our results suggest that, as age decreases, the contribution of phonation decreases, while that of plosive voicing increases. The perceptual down-weighting of phonation is in line with the direction of the tonal split process. We would expect the same trend for plosive voicing, but our results suggest the opposite direction. One possible explanation is the higher proficiency of younger Tamang speakers in Nepali, in which plosives have a phonological voicing contrast. Tamang villagers of 40 years ago did not distinguish /b, d, g/ from /p, t, k/ when speaking Nepali. The saliency of the voicing feature in Nepali, acquired in their second language, was likely transferred by younger speakers to their perception of prevoicing in their native language. If this is true, it would suggest that a contact-induced sound change, independent from the tonal split, is taking place in parallel (see Pearce [2009] for a similar sociolinguistic situation, due to contact with French, among town-dwelling male speakers of Kera). A comprehensive sociolinguistic study would be needed to tease apart language-internal and contact-induced factors.

4.4 Concluding remarks

In conclusion, we have shown that in Risiangku Tamang, newly emerged and residual cues from the tonal split process are all used in perception and enter in interaction with other cues in a stable variation. This situation is frequently encountered in voice register languages. Our data shows that it also exists in a tonal language after the establishment of a new tonal contrast. Whether or not Risiangku Tamang eventually evolves towards the disappearance of all residual cues, its present transitional state can last a long time and be made more complex by language-external factors such as contact.

Acknowledgments: We thank our Tamang consultants for their patient and enthusiastic collaboration, Christian Dicanio (Associate Editor) and two anonymous reviewers for valuable suggestions, Nicolas Audibert for his Praat script for resynthesizing prevoicing, the late Isabelle Tellier for her patient explanations on decision tree learning methods, Boyd Michailovsky for helpful comments, Amrit Yoncan for recording the instructions in Nepali, and Claire Michailovsky for drawing the pictures used for elicitation. Preliminary results have been reported at Atelier de Phonologie 2017, the Edinburgh Symposium on Historical Phonology 2017, Seoul International Conference on Speech Sciences 2017, Linguistic Society of Nepal 2018, and LabPhon 2018. We thank the audiences for their feedback. This study was conducted under the research strand "Evolutionary approaches to phonology", partly supported by a public grant overseen by the French National Research Agency (ANR) as part of the program "Investissements d'Avenir" (reference: ANR-10-LABX-0083 – LabEx EFL). It contributes to the IdEx Université de Paris – ANR-18-IDEX-0001.

References

Birkholz, Peter. 2013. Modeling consonant-vowel coarticulation for articulatory speech synthesis. *PLoS ONE* 8(4). e60603.

Boersma, Paul & David Weenink. 2016. *Praat: Doing phonetics by computer [Computer program]*. Version 6.0.21. Available at: http://www.praat.org/.

Brunelle, Marc & James Kirby. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass* 10(4). 191–207.

Brunelle, Marc, Ta Thành Tan, James Kirby & Đinh Lư Giang. 2020. Transphonologization of voicing in Chru: Studies in production and perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1). 15.

Coetzee, Andries W. 2017. Personal communication.

Coetzee, Andries W., Patrice Speeter Beddor, Kerby Shedden, Will Styler & Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66. 185–216.

Ferlus, Michel. 1979. Formation des registres et mutations consonantiques dans les langues Mon-Khmer [Development of phonation-type registers and consonant shifts in Mon-Khmer languages]. *Mon-Khmer Studies* 8. 1–76.

Gao, Jiayin. 2015. *Interdependence between tones, segments, and phonation types in Shanghai Chinese: Acoustics, articulation, perception, and evolution*. Paris: Sorbonne Paris Cité PhD Thesis. http://www.afcp-parole.org/doc/theses/these_JG15.pdf.

Gao, Jiayin & Takayuki Arai. 2019. Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics* 77. 1–33.

Gao, Jiayin, Pierre Hallé & Christoph Draxler. 2020. Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception? *Language and Speech* 63(3). 582–607.

Gao, Jiayin & Martine Mazaudon. 2017. *On the retention of an old feature in the Tamang dialect of Taglung (Tibeto-Burman, Nepal)*. HAL ID = halshs-01730119, version 1. Poster presented at the 4th Workshop on Sound Change, University of Edinburgh, 20–22 April 2017. Available at: https://halshs.archives-ouvertes.fr/halshs-01730119/file/WSC4_JGMM_poster.pdf.

Gao, Jiayin, Jihyeon Yun & Takayuki Arai. 2019. VOT-F0 coarticulation in Japanese: Production-biased or misparsing? In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International congress of phonetic sciences*, paper 619, 1–5. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Haudricourt, André-Georges. 1949. La conservation de la sonorité des sonores du thai commun dans le parler thô de Cao-bang [The preservation of the voicing of Common Thai (proto-Thai) voiced stops in the Tho dialect of Cao-bang]. In *Actes du XXIe Congrès International des Orientalistes*, 251–252. Paris: Société Asiatique.

Haudricourt, André-Georges. 1961. Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême-Orient [Two-way and three-way splits of tonal systems in some Far Eastern languages]. *Bulletin de la Société de Linguistique de Paris* 56(1). 163–180.

Haudricourt, André-Georges. 1965. Les mutations consonantiques des occlusives initiales en môn-khmer [Consonant shifts in Mon-Khmer initial stops]. *Bulletin de la Société de Linguistique de Paris* 60(1). 160–172.

Hildebrandt, Kristine A. 2005. A phonetic analysis of Manange segmental and suprasegmental properties. *Linguistics of the Tibeto-Burman Area* 28(1). 1–36.

Howe, Penelope Jane. 2017. *Tonogenesis in central dialects of Malagasy: Acoustic and perceptual evidence with implications for synchronic mechanisms of sound change*. Houston: Rice University PhD thesis.

Huffman, Franklin E. 1976. The register problem in fifteen Mon-Khmer languages. *Oceanic Linguistics Special Publications* 13. 575–589.

Hyslop, Gwendolyn. 2009. Kurtöp tone: A tonogenetic case study. *Lingua* 119. 827–845.

Idemaru, Kaori, Lori L. Holt & Seltman Howard. 2012. Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America* 132(6). 3950–3964.

Jiang, Bing'er, Meghan Clayards & Morgan Sonderegger. 2020. Individual and dialect differences in perceiving multiple cues: A tonal register contrast in two Chinese Wu dialects. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1). 11.

Kirby, James & Morgan Sonderegger. 2018. Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics* 70. 70–85.

Kuang, Jianjing & Mark Liberman. 2018. Integrating voice quality cues in the pitch perception of speech and non-speech utterances. *Frontiers in Psychology* 9. 1–11.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune Haubo Bojesen Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.

Lama, Santabir. 1959. *Syebu-syemu hvai-rimthim (Tamang jatiko vivahko riti thiti ra git) [Marriage songs and rituals of the Tamang]*. Darjeeling: Author.

Lenth, Russell. 2019. Emmeans: Estimated marginal means, aka least-squares means. Available at: https://CRAN.R-project.org/package=emmeans.

Maddieson, Ian & Peter Ladefoged. 1985. "Tense" and "lax" in four minority languages of China. *Journal of Phonetics* 13(4). 433–454.

Maspero, Henri. 1912. Études sur la phonétique historique de la langue annamite: Les initiales [Studies in Annamese historical phonetics: Initial consonants]. *Bulletin de l'École Française d'Extrême-Orient* 12. 1–124.

Mazaudon, Martine. 1973. *Phonologie du Tamang [Tamang phonology]*. Paris: Société d'études linguistiques et anthropologiques de France.

Mazaudon, Martine. 1978. Consonantal mutation and tonal split in the Tamang subfamily of Tibeto-Burman. *Kailash* 6(3). 157–179.

Mazaudon, Martine. 2005. On tone in Tamang and neighbouring languages: Synchrony and diachrony. In Shigeki Kaji (ed.), *Proceedings of the symposium cross-linguistic studies of tonal phenomena: Historical development, phonetics of tone and descriptive studies*, 79–96. Tokyo, Japan: ILCAA, Tokyo University of Foreign Studies.

Mazaudon, Martine. 2012. Paths to tone in the Tamang branch of Tibeto-Burman (Nepal). In Gunther De Vogelaer & Seiler Guido (eds.), *The dialect laboratory: Dialects as a testing ground for theories of language change* (Studies in Language Companion Series 128), 139–177. Amsterdam: John Benjamins.

Mazaudon, Martine & Alexis Michaud. 2008. Tonal contrasts and initial consonants: A case study of Tamang, a "missing link" in tonogenesis. *Phonetica* 65(4). 231–256.

Michailovsky, Boyd & Martine Mazaudon. 1994. Preliminary notes on the languages of the Bumthang group. In Per Kvaerne (ed.), *Tibetan studies (Proceedings of the 6th seminar of the International Association of Tibetan Studies)*, 545–557. Oslo, Norway: The Institute of Comparative Research in Human Culture.

Michaud, Alexis & Martine Mazaudon. 2006. Pitch and voice quality characteristics of the lexical word-tones of Tamang, as compared with level tones (Naxi data) and pitchplus- voice-quality tones (Vietnamese data). In Rüdiger Hoffmann & Hansjörg Mixdorff (eds.), *Proceedings of Speech Prosody 2006*, 819–822. Dresden, Germany: Technische Universität Darmstadt Press.

Milborrow, Stephen. 2019. *"rpart.plot" package*. Available at: https://cran.r-project.org/web/packages=rpart.plot.

Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary Frances Miller (eds.), *17th Regional meeting of the Chicago Linguistic Society: Papers from the parasession on language and behavior*, 178–203. Chicago, IL: Chicago Linguistic Society.

Pearce, Mary. 2009. Kera tone and voicing interaction. *Lingua* 119(6). 846–864.

Peirce, Jonathan & Michael MacAskill. 2018. *Building experiments in PsychoPy*. London: Sage.

Peralta, William. 2018. Tonogenesis: The perception of tone and the role of place of articulation in Kurtöp. In *Proceedings of the 6th international symposium on tonal aspects of languages*, 83–87. Berlin, Germany: Beuth Hochschule für Technik Berlin.

Pike, Kenneth L. 1970. The role of nuclei of feet in the analysis of tone in Tibeto-Burman languages of Nepal. In Hale Austin & Kenneth L. Pike (eds.), *Tone systems of Tibeto-Burman languages of Nepal*, 37–48. Urbana: Department of Linguistics, University of Illinois.

Pinget, Anne-France, René Kager & Hans Van de Velde. 2020. Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech* 63(3). 660–685.

Pittayaporn, Pittayawat & James Kirby. 2016. Laryngeal contrasts in the Tai dialect of Cao Bằng. *Journal of the International Phonetic Association* 47(1). 65–85.

R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Schertz, Jessamyn & Emily J. Clare. 2020. Phonetic cue weighting in perception and production. *WIREs Cognitive Science* 11(2). e1521.

Therneau, Terry, Beth Atkinson & Brian Ripley. 2019. *"rpart" package*. Available at: https://cran.rproject.org/web/packages=rpart.

Tian, Jia & Jianjing Kuang. 2021. The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association* 51(2). 202–228.

Valbret, Hélène, Eric Moulines & Jean-Pierre Tubach. 1992. Voice transformation using PSOLA technique. *Speech Communication* 11(2–3). 175–187.

Yang, Cathryn, James N. Stanford & Zhengyu Yang. 2015. A sociotonetic study of Lalo tone split in progress. *Asia-Pacific Language Variation* 1(1). 52–77.

Zhang, Jie & Hanbo Yan. 2018. Contextually dependent cue realization and cue weighting for a laryngeal contrast in Shanghai Wu. *The Journal of the Acoustical Society of America* 144(3). 1293–1308.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/lingvan-2021-0085).